Monitoring Markov Chains

Mahyar Karimi

Institute of Science and Technology Austria

September 13, 2023

Coins

- Given two (*maybe* fair) coins:

- How different are the coins?

Coins

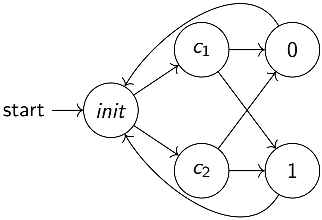
- Given two (*maybe* fair) coins:

- How different are the coins?

$$\mathbb{P}\left(1 \mid c_1\right) - \mathbb{P}\left(1 \mid c_2\right)$$

Coins in Markov Chains

- Let's turn this into a Markov chain:

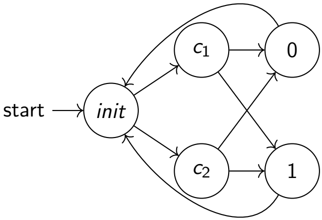


- Observing this Markov chain $\rightarrow$ Stream of states:

$$init, c_1, 1, init, c_2, 0, init, c_1, 0, \ldots$$

Coins in Markov Chains

- Let's turn this into a Markov chain:



- Observing this Markov chain $\rightarrow$ Stream of states:

$$init, c_1, 1, init, c_2, 0, init, c_1, 0, \dots$$

- How to estimate $\mathbb{P}(1 \mid c_1) - \mathbb{P}(1 \mid c_2)$?

What We Developed

- Techniques for estimating expressions over $\mathbb{P}(\cdot \mid \cdot)$
  - Frequentist and Bayesian: two well-known paradigms in statistics

What We Developed

- Techniques for estimating expressions over $\mathbb{P}(\cdot \mid \cdot)$
  - Frequentist and Bayesian: two well-known paradigms in statistics
- $\implies$ Monitors for transitions in Markov chains

What We Developed

- Techniques for estimating expressions over $\mathbb{P}(\cdot \mid \cdot)$
  - Frequentist and Bayesian: two well-known paradigms in statistics
- $\implies$ Monitors for transitions in Markov chains
- $\implies$ Monitoring fairness properties!

The Frequentist, The Bayesian

- The *frequentist*
  - Probabilities $\equiv$ Long-run frequencies
  - Estimating the *ground truth*.

The Frequentist, The Bayesian

- The *frequentist*
  - Probabilities $\equiv$ Long-run frequencies
  - Estimating the *ground truth*.
- The *Bayesian*
  - Markov chains sampled from a given distribution (*prior belief*).
  - **No ground truth!** Only prior belief.

Frequentist Monitor

- Monitor: Input stream $\rightarrow$ Inteval stream:

$$\mathcal{A} : \Sigma^* \rightarrow I_{-\infty,\infty}$$

[1] **P**robably **A**pproximately **C**orrect

Frequentist Monitor

- Monitor: Input stream $\rightarrow$ Inteval stream:

$$\mathcal{A} : \Sigma^* \rightarrow I_{-\infty,\infty}$$

- The monitor gives PAC[1]-style guarantees (for given δ):

$$\mathbb{P}\Big(\varphi \in \mathcal{A}(\vec{X})\Big) \geq 1 - \delta$$

[1] **P**robably **A**pproximately **C**orrect

Frequentist Monitor

- Monitor: Input stream $\rightarrow$ Inteval stream:

$$\mathcal{A} : \Sigma^* \rightarrow I_{-\infty,\infty}$$

- The monitor gives PAC[1]-style guarantees (for given δ):

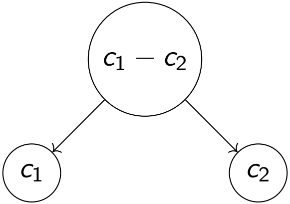
$$\mathbb{P}\Big(\varphi \in \mathcal{A}(\vec{X})\Big) \geq 1 - \delta$$

- Intervals $\Longleftarrow$ Bounding with Hoeffding's inequality.

[1] **P**robably **A**pproximately **C**orrect

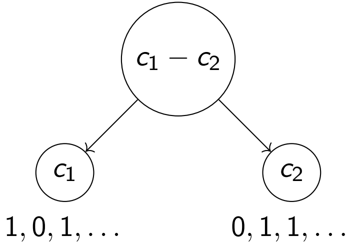
Frequentist and Coins

- For $\mathbb{P}(1 \mid c_1) - \mathbb{P}(1 \mid c_2)$:



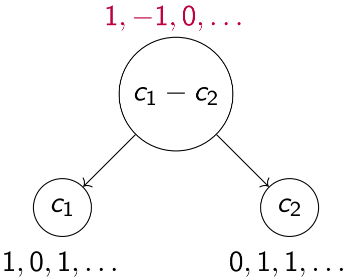
Frequentist and Coins

- For $\mathbb{P}(1 \mid c_1) - \mathbb{P}(1 \mid c_2)$:



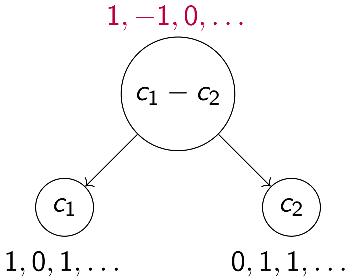
Frequentist and Coins

- For $\mathbb{P}(1 \mid c_1) - \mathbb{P}(1 \mid c_2)$:



Frequentist and Coins

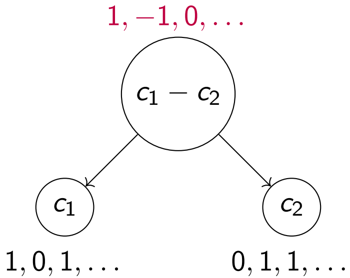
- For $\mathbb{P}(1 \mid c_1) - \mathbb{P}(1 \mid c_2)$:



- New RV with the **same expected value** as the subtraction

Frequentist and Coins

- For $\mathbb{P}(1 \mid c_1) - \mathbb{P}(1 \mid c_2)$:



- New RV with the **same expected value** as the subtraction
- A few other techniques for multiplication and division

The Bayesian

- Fixed *prior distribution*[2] θ over Markov chains *(prior belief)*
- Obtaining *posterior belief* after observing $\vec{x}$

[2] Matrix beta distribution

Bayesian Monitor

- The monitor guarantees the following (for given δ, θ):

$$\forall \vec{x} \in \Sigma^*. \; \mathbb{P}_\theta\Big(\varphi(\mathcal{M}) \in \mathcal{A}_\theta(\vec{x})\Big) \geq 1 - \delta$$

Bayesian Monitor

- The monitor guarantees the following (for given δ, θ):

$$\forall \vec{x} \in \Sigma^*.\ \mathbb{P}_\theta\Big(\varphi(\mathcal{M}) \in \mathcal{A}_\theta(\vec{x})\Big) \geq 1 - \delta$$

- Core component: efficient, incremental computation of $\mathbb{E}_\theta\left(\varphi(\mathcal{M}) \mid \vec{x}\right)$:

$$\mathbb{E}_\theta\left(\varphi(\mathcal{M}) \mid \vec{x}ab\right) = \mathbb{E}_\theta\left(\varphi(\mathcal{M}) \mid \vec{x}a\right) \cdot \frac{c_{ab}(\vec{x}) + d_{ab}}{c_{ab}(\vec{x})} \cdot \frac{c_a(\vec{x})}{c_a(\vec{x}) + d_a}$$

Base case comes from θ.

Bayesian Monitor

- The monitor guarantees the following (for given δ, θ):

$$\forall \vec{x} \in \Sigma^*.\ \mathbb{P}_\theta\Big(\varphi(\mathcal{M}) \in \mathcal{A}_\theta(\vec{x})\Big) \geq 1 - \delta$$

- Core component: efficient, incremental computation of $\mathbb{E}_\theta\left(\varphi(\mathcal{M}) \mid \vec{x}\right)$:

$$\mathbb{E}_\theta\left(\varphi(\mathcal{M}) \mid \vec{x}ab\right) = \mathbb{E}_\theta\left(\varphi(\mathcal{M}) \mid \vec{x}a\right) \cdot \frac{c_{ab}(\vec{x}) + d_{ab}}{c_{ab}(\vec{x})} \cdot \frac{c_a(\vec{x})}{c_a(\vec{x}) + d_a}$$

  Base case comes from θ.

- Intervals $\Longleftarrow$ Bounding with Chebyshev's inequality.

Monitoring Regular Expressions?

- So far: $\mathbb{P}\left(b \mid a\right)$

Monitoring Regular Expressions?

- So far: $\mathbb{P}\left(b \mid a\right)$
- $\mathbb{P}\left(b_1 \dots b_k \mid a\right)$ is not that different:

$$\mathbb{P}\left(b_1 \dots b_k \mid a\right) = \mathbb{P}\left(b_1 \mid a\right) \cdots \mathbb{P}\left(b_k \mid b_{k-1}\right)$$

- $\rightarrow^*$ Monitoring *star-free* regular expression r: $\mathbb{P}\left(r \mid a\right)$.